APPROXIMATE EVALUATION OF FIRST-ORDER COUNTING QUERIES

Jan Dreier, Peter Rossmanith

January 12, 2021

RWTH Aachen University

SODA 2021

1

MSO on treewidth



MSO on treewidth FO on sparse graphs









 \bigcirc independent set of size k:

$$\exists x_1 \dots \exists x_k \bigwedge_{i,j} x_i \not\sim x_j \land x_i \neq x_j$$

 \bigcirc independent set of size k:

$$\exists x_1 \dots \exists x_k \bigwedge_{i,j} x_i \not\sim x_j \land x_i \neq x_j$$

 \bigcirc dominating set of size k:

$$\exists x_1 \dots \exists x_k \, \forall y \, \bigvee_i y \sim x_i \lor y = x_i$$

 \bigcirc independent set of size k:

$$\exists x_1 \dots \exists x_k \bigwedge_{i,j} x_i \not\sim x_j \land x_i \neq x_j$$

 \bigcirc dominating set of size k:

$$\exists x_1 \dots \exists x_k \, \forall y \, \bigvee_i y \sim x_i \lor y = x_i$$

○ basic database queries

 \bigcirc independent set of size k:

$$\exists x_1 \dots \exists x_k \bigwedge_{i,j} x_i \not\sim x_j \land x_i \neq x_j$$

 \bigcirc dominating set of size k:

$$\exists x_1 \dots \exists x_k \, \forall y \, \bigvee_i y \sim x_i \lor y = x_i$$

○ basic database queries

Best algorithms on general graphs: $n^{O(k)}$

Model-Checking



Model-Checking



$MC(\mathcal{G}, L)$

Input: A graph $G \in \mathcal{G}$ and a sentence $\varphi \in L$

Parameter: $|\varphi|$

Problem: Is φ true in G?

Goal: linear FPT run time $f(|\varphi|)n$

Model-Checking



MC(\mathcal{G} , L)Input:A graph $G \in \mathcal{G}$ and a sentence $\varphi \in L$ Parameter: $|\varphi|$ Problem:Is φ true in G?

Goal: linear FPT run time $f(|\varphi|)n$



If \mathcal{G} has bounded treewidth then MC(\mathcal{G} , MSO) \in FPT.

[Courcelle 1990]



If \mathcal{G} has bounded treewidth then MC(\mathcal{G} , MSO) \in FPT.

[Courcelle 1990]



If \mathcal{G} has bounded treewidth then MC(\mathcal{G} , MSO) \in FPT.

[Courcelle 1990]



If \mathcal{G} has bounded treewidth then MC(\mathcal{G} , MSO) \in FPT.

[Courcelle 1990]

If ${\mathcal G}$ is nowhere dense then MC(${\mathcal G},$ FO) \in FPT.

[Grohe, Kreutzer, Sieberz 2011]



If \mathcal{G} has bounded treewidth then MC(\mathcal{G} , MSO) \in FPT.

[Courcelle 1990]

If ${\mathcal G}$ is nowhere dense then MC(${\mathcal G},$ FO) \in FPT.

[Grohe, Kreutzer, Sieberz 2011]



|E| $\leq c$

for every graph in the graph class.



|E| ≤ 2

for every graph in the graph class.









$$\frac{|E|}{|V|} \le f(r)$$

for every r-shallow minor of every graph in the graph class.

Partial Dominating Set		
Input:	A graph G and $k,m\in {f N}$	
Parameter:	k	
Problem:	Are there k vertices dominating m vertices?	

Partial Dominating Set		
Input:	A graph G and $k,m\in {f N}$	
Parameter:	k	
Problem:	Are there \boldsymbol{k} vertices dominating \boldsymbol{m} vertices?	

Cannot be expressed in first-order logic (requires $\exists y_1 \dots \exists y_m$).

Partial Dominating Set		
Input:	A graph G and $k,m \in \mathbf{N}$	
Parameter:	k	
Problem:	Are there k vertices dominating m vertices?	

Cannot be expressed in first-order logic (requires $\exists y_1 \dots \exists y_m$).

Can be solved on H-minor free graphs in time $(g(H)k)^k n^{O(1)}.$ [Amini, Fomin, Saurabh, 2008]

Can be solved on apex-minor-free graphs in time $2^{\sqrt{k}}n^{O(1)}$. [Fomin, Lokshtanov, Raman, Saurabh, 2011]

Is W[1]-hard for 2-degenerate graphs. [Golovach, Villanger 2008]

Partial Dominating Set		
Input:	A graph G and $k, m \in \mathbf{N}$	
Parameter:	k	
Problem:	Are there \boldsymbol{k} vertices dominating \boldsymbol{m} vertices?	

 $FO(\{>0\}) = FO + "there are at least/most <math>m \in \mathbb{N}$ elements"

Partial Dominating Set		
Input:	A graph G and $k,m\in \mathbf{N}$	
Parameter:	k	
Problem:	Are there k vertices dominating m vertices?	

 $FO(\{>0\}) = FO + "there are at least/most <math>m \in \mathbb{N}$ elements"

$$\exists x_1 \dots \exists x_k \# y \left(\bigvee_i y \sim x_i \land y = x_i\right) \ge m$$

Partial Dominating Set		
Input:	A graph G and $k,m\in \mathbf{N}$	
Parameter:	k	
Problem:	Are there k vertices dominating m vertices?	

 $FO(\{>0\}) = FO + "there are at least/most <math>m \in \mathbb{N}$ elements"

$$\exists x_1 \dots \exists x_k \, \# y \, (\bigvee_i y \sim x_i \wedge y = x_i) \ge m$$

Length of formula depends only on k (and not on m)

Definition of $FO(\{>0\})$

built recursively using

- the rules of FO
- $\circ \ \# y \ \varphi \geq m$ for every $m \in \mathbb{N}$ and FO($\{>0\}$) formula φ

Example 1: PARTIAL DOMINATING SET

$$\exists x_1 \dots \exists x_k \, \# y \, \left(\bigvee_i y \sim x_i \land y = x_i\right) \ge m$$

Example 2: *h*-Index

#mypaper (#otherpaper cite(otherpaper, mypaper) $\ge h$) $\ge h$



If $\mathcal G$ has bounded degree then MC($\mathcal G$, FOC) \in FPT. [Kuske, Schweikardt 2017]



If \mathcal{G} has bounded degree then MC(\mathcal{G} , FOC) \in FPT. [Kuske, Schweikardt 2017]

 $MC(\mathcal{G}, FO(\{>0\}))$ is AW[*]-hard on trees.

similar to [Grohe, Schweikardt 2018]

Bad News



 \Leftrightarrow





satisfies FO({>0}) formula

Bad News





 \Leftrightarrow



satisfies FO({>0}) formula



Bad News


Bad News



Are there k vertices dominating at least m = 5000 vertices?



Are there k vertices dominating at least m = 4983 vertices?



Are there k vertices dominating at least m = 5017 vertices?



Are there k vertices dominating at least m = 5017 vertices?



Let $\varepsilon > 0$. A formula φ is ε -unstable on a graph G if scaling the counting literals by $(1 \pm \varepsilon)$ changes whether φ is true in G.

Let \mathcal{G} be a graph class with bounded expansion and $\varepsilon > 0$.

Let \mathcal{G} be a graph class with bounded expansion and $\varepsilon > 0$. There exists an algorithm which takes $G \in \mathcal{G}$, $\varphi \in FO(\{>0\})$, runs in time $f(|\varphi|)n$ and returns (\circ, \circ) , (\circ, \circ) .

Let \mathcal{G} be a graph class with bounded expansion and $\varepsilon > 0$. There exists an algorithm which takes $G \in \mathcal{G}$, $\varphi \in FO(\{>0\})$, runs in time $f(|\varphi|)n$ and returns (\circ, \circ) , or (\circ, \circ) .



Let \mathcal{G} be a graph class with bounded expansion and $\varepsilon > 0$. There exists an algorithm which takes $G \in \mathcal{G}$, $\varphi \in FO(\{>0\})$, runs in time $f(|\varphi|)n$ and returns (\circ, \circ) , (\circ, \circ) .





Let \mathcal{G} be a graph class with bounded expansion and $\varepsilon > 0$. There exists an algorithm which takes $G \in \mathcal{G}$, $\varphi \in FO(\{>0\})$, runs in time $f(|\varphi|)n$ and returns (e) , (e) , or (e) .





- \bigcirc If $(\ref{eq: eq: false})$ then φ is false on G.
- \bigcirc If $(\bullet \bullet)$ then φ is ε -unstable on G.



Partial Dominating Set: $\exists x_1 \dots \exists x_k \# y (\bigvee_i y \sim x_i \land y = x_i) \ge m$

There exists a set dominating $\geq (1 + \varepsilon)m$ vertices.



 $x_1 \dots x_k$

 $x_1 \dots x_k$

 $x_1 \dots x_k$



There exists a set dominating $\geq (1 + \varepsilon)m$ vertices.





There exists a set dominating $\geq (1 + \varepsilon)m$ vertices.

All sets dominate $< (1 + \varepsilon)m$ vertices and there exists a set dominating $\ge (1 - \varepsilon)m$ vertices.



 $x_1 \dots x_k$

 $x_1 \dots x_k$

All sets dominate $< (1 - \varepsilon)m$ vertices.

PARTIALDOMINATINGSET can be solved in time f(k)n on graph classes with bounded expansion.

PARTIALDOMINATINGSET can be solved in time f(k)n on graph classes with bounded expansion.

This holds for all problems of the form

$$\exists x_1 \dots \exists x_k \# y \ \underbrace{\varphi(yx_1 \dots x_k)}_{\in \mathrm{FO}}.$$

Theorem

Approximate model-checking becomes hard on trees if also allow one of the following:

Theorem

Approximate model-checking becomes hard on trees if also allow one of the following:

 \bigcirc comparing #y and #z

(e.g., $\#y \ \varphi > \#z \ \psi$)

Theorem

Approximate model-checking becomes hard on trees if also allow one of the following:

- \bigcirc comparing #y and #z
- \bigcirc counting tuples #yz

(e.g., $\#y \varphi > \#z \psi$) (e.g., $\#yz \varphi > m$)

Theorem

Approximate model-checking becomes hard on trees if also allow one of the following:

- \bigcirc comparing #y and #z
- \bigcirc counting tuples #yz
- multiplying of counting terms

(e.g., $\#y \ \varphi > \#z \ \psi$)

- (e.g., $\#yz \; arphi > m$)
- (e.g., $\#y \ \varphi \cdot \#z \ \psi > m$)

Theorem

Approximate model-checking becomes hard on trees if also allow one of the following:

- \bigcirc comparing #y and #z
- \bigcirc counting tuples #yz
- multiplying of counting terms
- subtraction of counting terms

(e.g., $\#y \ \varphi > \#z \ \psi$)

- (e.g., $\#yz \ \varphi > m$)
- (e.g., $\#y \ \varphi \cdot \#z \ \psi > m$)
- (e.g., $\#y \ \varphi \#z \ \psi > m$)



FO($\{>0\}$) is

○ hard to solve exactly on trees,



FO($\{>0\}$) is

- hard to solve exactly on trees,
- possible to approximate on bounded expansion.

FO($\{>0\}$) is

- hard to solve exactly on trees,
- possible to approximate on bounded expansion.

Slight extensions of FO($\{>0\}$) are

○ hard to approximate on trees.

FO($\{>\!0\}$) is

- hard to solve exactly on trees,
- possible to approximate on bounded expansion.

Slight extensions of FO($\{>0\}$) are

○ hard to approximate on trees.

 \Rightarrow FO({>0}) seems like "the right logic" for approximation on sparse graphs



Can we generalize our results to nowhere dense graph classes?

$$m_1 \le \# x_1 \Big($$

We want to gradually simplify this formula.

$$m_1 \le \# x_1 \left(m_2 \le \# x_2 \right)$$

))

$$m_1 \le \# x_1 \left(m_2 \le \# x_2 \left(m_3 \le \# x_3 \right) \right)$$

$$m_1 \leq \#x_1 \left(m_2 \leq \#x_2 \left(m_3 \leq \#x_3 \quad \overbrace{\varphi(x_1 x_2 x_3)}^{\text{quantifer-free FO}} \right) \right)$$

$$m_1 \leq \#x_1 \left(m_2 \leq \#x_2 \left(\underbrace{m_3 \leq \#x_3}_{\text{replace with quantifier-free FO}} \varphi(x_1 x_2 x_3) \right) \right)$$

$$m_1 \leq \#x_1 \left(m_2 \leq \#x_2 \quad \overbrace{\varphi'(x_1 x_2)}^{\text{quantifier-free FO}} \right)$$

We want to gradually simplify this formula.

quantifier-free FO

$$m_1 \le \# x_1 \left(\underbrace{m_2 \le \# x_2}_{\text{replace with quaptifier-free FO}} \phi'(x_1 x_2) \right)$$

replace with quantifier-free FO

We want to gradually simplify this formula.

 $\underset{m_1 \leq \# x_1}{\operatorname{quantifier-free FO}} \widetilde{\varphi''(x_1)}$

We want to gradually simplify this formula.

quantifier-free FO $m_1 \leq \# x_1$ x_1

replace with quantifier-free FO


$$\underbrace{\#x_3 \ (x_3 \sim x_1 \lor x_3 \sim x_2) \ge m}_{\text{replace with quantifier free EQ}}$$

replace with quantifier-free FO











 $\begin{array}{l} R_{\geq i}(x) \text{ true} \\ \text{iff } |N(x)| \geq i \end{array}$





$$\bigvee_{i=0}^{1/\varepsilon} R_{\geq \varepsilon mi}(x_1) \wedge R_{\geq m-\varepsilon mi}(x_2)$$



$$\bigvee_{i=0}^{1/\varepsilon} R_{\geq \varepsilon mi}(x_1) \wedge R_{\geq m-\varepsilon mi}(x_2)$$



$$\bigvee_{i=0}^{1/\varepsilon} R_{\geq \varepsilon mi}(x_1) \wedge R_{\geq m-\varepsilon mi}(x_2)$$



$$\bigvee_{i=0}^{1/\varepsilon} R_{\geq \varepsilon mi}(x_1) \wedge R_{\geq m-\varepsilon mi}(x_2)$$



$$\bigvee_{i=0}^{1/\varepsilon} R_{\geq\varepsilon mi}(x_1) \wedge R_{\geq m-\varepsilon mi}(x_2)$$

$$\varepsilon m \quad 2\varepsilon m \quad 4\varepsilon m \quad 5\varepsilon m \quad 6\varepsilon m \quad 7\varepsilon m \cdots m$$



$$\bigvee_{i=0}^{1/\varepsilon} R_{\geq \varepsilon mi}(x_1) \wedge R_{\geq m-\varepsilon mi}(x_2)$$



$$\bigvee_{i=0}^{1/\varepsilon} R_{\geq \varepsilon m i}(x_1) \wedge R_{\geq m-\varepsilon m i}(x_2)$$













We assume (for simplicity) x_1 has only one x_2 with a large intersection.



We assume (for simplicity) x_1 has only one x_2 with a large intersection. We call it $f(x_1)$.



 $\begin{array}{c} Q_f(x) \text{ true} \\ \text{iff } |N(x) \cup N(f(x))| \geq m \end{array}$



 $\begin{array}{c} Q_f(x) \text{ true} \\ \text{iff } |N(x) \cup N(f(x))| \geq m \end{array}$

Final Formula:

$$\left(x_2 = f(x_1) \land Q_f(x_1)\right)$$



 $\begin{array}{c} Q_f(x) \text{ true} \\ \text{iff } |N(x) \cup N(f(x))| \geq m \end{array}$

Final Formula: $\left(x_2 = f(x_1) \land Q_f(x_1) \right) \lor$ $\left(x_2 \neq f(x_1) \land \varphi_{\text{small}}(x_1, x_2) \right)$

$$m_1 \leq \#x_1 \left(m_2 \leq \#x_2 \left(\underbrace{m_3 \leq \#x_3}_{\text{replace with quantifier-free FO}} \right) \right)$$

quantifier-free FO

$$m_1 \le \# x_1 \left(m_2 \le \# x_2 \quad \overbrace{\varphi'(x_1 x_2)}^{\text{quantifier-free FO}} \right)$$

$$m_1 \leq \#x_1 \left(\underbrace{m_2 \leq \#x_2}_{\text{replace with quantifier-free FO}} \varphi'(x_1 x_2) \right)$$

quantifier-free FO $m_1 \le \# x_1$ $\varphi''(x_1)$

quantifier-free FO $m_1 \le \# x_1$ (x_1) replace with guantifier-free FO

Gradually simplify formula.

quantifier-free FO



Questions or Feedback?

dreier@cs.rwth-aachen.de